



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 134765

TO: Jon E Angell  
Location: rem/2d20/2c18  
Art Unit: 1635  
Friday, October 08, 2004

Case Serial Number: 10/061201

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen 1A69  
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

### Search Notes

O'Bryen, Barbara

---

**From:** Angell, Jon E  
**Sent:** Tuesday, October 05, 2004 3:35 PM  
**To:** O'Bryen, Barbara  
**Subject:** Sequence Database Search Request

SEARCH REQUEST FORM  
Scientific and Technical Information Center

Examiner# : 78697  
Art Unit : 1635  
Phone Number: 571-272-0756  
Date: 10/05/04  
Serial Number: 10/061,201  
Mailbox & Bldg/Room Location: REMSEN 2C18  
Results Format Preferred (circle): **DISK**

I would like to have a search of the standard databases performed using the following SEQ. ID NOs. from application :  
10/061,201

SEQ ID NO. 3

NOTE: SEQ ID NO: 3 is an amino acid sequence, but what I need is a search that identifies any NUCLEIC ACID SEQUENCE that encodes the amino acid sequence that is SEQ ID NO: 3.

Please search for the amino acid sequence of SEQ ID NO: 3 AND  
any nucleic acid that encodes the amino acid sequence of SEQ ID NO. 3

Thanks, Let me know if you need any further clarification...  
Eric

*J. Eric Angell*  
Art Unit 1635  
Office: REMSEN 2D20  
mailbox: REM 2C18  
571-272-0756

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 7, 2004, 08:28:49; Search time 6448 Seconds  
(without alignments)  
4900.290 Million cell updates/sec

Title: US-10-061-201-3

Perfect score: 3791

Sequence: 1 MDDLTLDDLLEPCVCFEKL.....AKATTLVSTASGTQTVFPSK 729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delect 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n model -DEV=xip  
-Q=cgnt2\_1USPTO\_spool\_US10061201/runat\_05102004\_174113\_9275/app\_query.fasta\_1\_903  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rgs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=prt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10061201 @CGN 1.1 5600 @runat\_05102004\_174113\_9275 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.hig.\*
- 3: gb.in.\*
- 4: gb.em.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sis.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*

- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sis.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.hig.hum.\*
- 31: em.hig.in.\*
- 32: em.hig.other.\*
- 33: em.hig.mus.\*
- 34: em.hig.pln.\*
- 35: em.hig.rod.\*
- 36: em.hig.nant.\*
- 37: em.hig.vt.\*
- 38: em.sy.\*
- 39: em.higo.hum.\*
- 40: em.higo.mus.\*
- 41: em.higo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result				Query				Description	
No.	Score	Match	Length	DB	ID				
1	3791	100.0	2190	6	AX530493				
2	3791	100.0	2497	6	AX530492				
3	3765	99.3	5608	9	HSM804610				
4	3762	99.2	3019	9	AK074234				
5	3219	84.9	1855	6	AX530497				
6	3219	84.9	2086	6	AX530495				
7	1446	38.1	1026	9	BC031650				
8	1445	38.1	1008	9	AK058046				
9	1167	30.8	2368	10	BC048400				
10	1067	28.1	2685	10	AF515735				
11	1049.5	27.7	2667	6	AX776591				
12	1049.5	27.7	5128	6	AX776593				
13	1042	27.5	3206	6	AX776598				
14	1042	27.5	3206	10	AF030131				
15	1041.5	27.5	3251	6	AX540752				
16	1033.5	27.3	2331	6	AX776594				
17	1033.5	27.3	2332	6	AX878405				
18	1033.5	27.3	2332	6	BD157200				
19	1033.5	27.3	2332	9	AK021429				
20	1006	26.5	3146	6	AX675069				
21	1004	26.5	750	6	AR413566				
22	1004	26.5	750	6	BD109119				
23	986.5	26.0	5668	9	BC053671				
24	914.5	24.1	5074	10	BC060113				
25	914.5	24.1	5074	10	BC060696				
26	839	22.1	677	6	BD217010				
27	756.5	20.0	1977	3	AK112836				
28	745.5	19.7	1672	9	BC057304				
29	745.5	19.7	1672	9	BC041023				
30	708.5	18.7	3149	3	AY051984				
31	708.5	18.7	3149	6	AX776600				
32	707.5	18.7	3123	3	AF220364				
33	677	17.9	195887	9	AC011359				
34	675	17.8	500	6	AX530510				
35	671	17.7	484	6	AX530500				
36	666.5	17.6	112088	9	AC005216				
37	655.5	17.3	171095	9	AC091887				
38	626	16.5	534	6	AR413567				
39	626	16.5	534	6	BD109120				
40	609.5	16.1	175506	3	AC008004				
41	609.5	16.1	194897	3	AC007697				
42	609.5	16.1	246095	3	AE003802				



Db 1441 CCCTTCAAATCCCGTCTTTGTGCCCACTGCCATAGTCAACCCCGTGTAGAGACACAGCCGGC 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1501 CCTGGGACTTTAGGACAAAGGTCTCTTCGAAAGGGCGGAGCAGCATGAGAAAGATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGACAGAGACCCCTCCAGTCCGGGATCCCCCACTCTCTGTTAGGCTCCCTCAGACGC 1620  
Qy 541 SerProThrMetValLeuArgProGlnPheGlnPheThrGlnProGlnGlyIlePro 560  
Db 1621 ACCCCACCATGGTCTCTTCCGCTCAGCAGTTCCAAATTTCTACCAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValGlnMetGlySerLysProAlaLeuThrGlyGlnPro 580  
Db 1681 TCCTCCCTCCTCAGCCGTGGTGGAGATGGGGTCCAAAGCTGCCCTCAGCGGGAGGCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGlnAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1741 GCCCTCAGTGCATCAGCAGGGGCGAGTGAGGCCCGGACCCCACTCCGGGGCCAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProLysSerGluProLysProLysProAlaSer 620  
Db 1801 ATTATGGAAGACAAAGAAATCCCCATCAAGAGTGAGGCTCTTGCCAAACCGCCGCACTC 1860  
Qy 621 AlaProSerLeuLeuValLysProGluAsnSerArgAsnGlyIleGlnLysGlnVal 640  
Db 1861 GCCCCACCATCCTCTGTTGTAACACAGAAAACTCAAGAAATGGCATCGAAGAAAGCATC 1920  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro 660  
Db 1921 AAAACCGTGAGATTTCAGAAATTACAGGCCCTCTCCACCAACAATTTACACCTCCCATCCC 1980  
Qy 661 ThrSerGlyLysProGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
Db 1981 ACCTCCGAAAGCCCTGAACAGCCAGCCACCCCTCAAGGGGTCCAGCCTGAAGCAGCGTCC 2040  
Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2041 TTGGGGCCAGAGATGACCGTCTTATTTGGCCACCCGAAAGTGCGCTGCCCACTCCGGACAGCAG 2100  
Qy 701 ThrAspLeuArgArgLysSerLeuAlaLysAlaThrThrLeuValSerThrAlaSer 720  
Db 2101 ACAGACCTCCGGAGAAAGTCAAGTCACTTGCCAAAGCCCAACCCCTGGTGTCTCACTGCCTCA 2160  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
Db 2161 GGCACGACGACCGTGTTCACGACAAA 2187

RESULT 2

AX530492

LOCUS AX530492 2497 bp DNA linear PAT 22-NOV-2002

DEFINITION Sequence 1 from Patent EP1239051.

ACCESSION AX530492

VERSION AX530492.1 GI:25252369

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,

Mammalia, Eutheria, Primates, Catarrhini, Homnidae, Homo.

REFERENCE 1

AUTHORS Shamon,M

TITLE Human posh-like protein 1

JOURNAL Patent: EP 1239051-A 1 11-SEP-2002;

Acemica, Inc. (US)

FEATURES

Location/Qualifiers

source 1..2497

/organism="Homo sapiens"

/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
Pred No.: 509c-149 Length: 2497  
Score: 379100 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-061-201-3 (1-729) x AX530492 (1-2497)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 232 ATGGATGATTTGACGCTTACTTGATCTCTTGGAGTGGCCCTGTGTGCTTTGAGAAGCTCGAT 291  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLysLeuGlnArgVal 40  
Db 292 GTCACAGCCAAAGTCTCCTTGGCCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 351  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 352 TTCAAGGCCCCACAAAGAGTGCGGTCGCCCGGAATGCAGGACGCGCTGTGTTTCCAAACATT 411  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 412 GAGCGCGTGCCTGCCGCCAACCTGCTGCTGCTGGATGGAGTGGCGCTCAGGGCAG 471  
Qy 81 SerSerArgGlyGlySerPheArgArgProGlyThrMetLeuGlnAspGlyArg 100  
Db 472 AGCTCCGGGAGAGGGGGCTCTTCCGCCAGGCTTGGACAGTACACCTTCGAGATGGCAGG 531  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 532 AAAAGCAGGACCAACCCAGACAGCTCTGACGGCCAGTCTTTCGGGCTAGTGCCTAATGTC 591  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 592 AGAATCCACATGGATGGGTGGCTCGAGCAAGGCCCTTATGCAACTACAGAGGGCAGAA 651  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 652 CCCGTCACCTAAGGTTTAAATAAGGGAGATATCATCTCTTCCGGAGACAGCTTGTATGAG 711  
Qy 161 AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 712 AATTGGTACCAGGGGGAAATCAATGGCATCAGCGGGAACCTTCCAGCCAGCTCCGTGGAA 771  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 772 GTCATCAAGCAGCTGCCCGCAGCGCCCGCTCTGCGAGGCCCTCTACAACCTTCGACCTA 831  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220  
Db 832 CGAGGCAAGGCAAGAGTGAACACAGGATTGGCTGACCTTCTCAAGGACCGATATCATC 891  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 892 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAGAAAGGCAAGTTACGAGATAAAGTAGGC 951  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 952 ATCTTCCCTATCTGTGTTGTAGAGCCCAAAOCTCACCCGCAAGACACCTTTTGTAGAGA 1011  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 1012 AAAAGGTCCCGAGTCACTCTGCACAAAAAACCTGTCTGCTGCTGCTCTCCAGAGGC 1071

Qy 621 AlaProSerLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
 Db 2092 CCCCCACCAATCCATCTGCTGGTGAACACGAAAACTCAAGAAATGGGATCGAAAGCAAGTC 2151  
 Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro 660  
 Db 2152 AAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACCAACAAACATTACACCTCCCATCC 2211  
 Qy 661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaSer 680  
 Db 2212 ACCTCCGGGAAAGCCTGAACAGCCAGCCACCTCAAGGCGTCCAGCGCTGAAGCAGCGTCC 2271  
 Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
 Db 2272 TTGGGCCACAGATGACCGCTCTATTTCGCCACCGAAGTGCGCTGCCACTCCGGACACGAC 2331  
 Qy 701 ThrAspLeuArgLysSerAlaLeuAlaLysAlaThrLeuValSerThrAlaSer 720  
 Db 2332 ACAGACCTCCGGGAAAGAACTGACGTCTTGCCAGGCCACACCTGGTGTCCACTGCGCTCA 2391  
 Qy 721 GlyThrGlnThrValPheProSerLys 729  
 Db 2392 GCGACGACGACCGTGTTCGCCAGCAAA 2418  
 RESULT 3  
 HSM804610  
 LOCUS HSM804610 5608 bp mRNA linear PRI 13-MAY-2003  
 DEFINITION Homo sapiens mRNA; cDNA DKFP451J159 (from clone DKFP451J159).  
 ACCESSION AL833297  
 VERSION AL833297.1 GI21733931  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 5608)  
 AUTHORS Otterwaelder B., Obermaier B., Deuschle B., Mewes H.W.,  
 Weil B., Arnold C., Osanger A., Fobo G., Han M. and Wiemann S.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr. 1, D-85764  
 Neuberg, GERMANY  
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by MediGenomix (Martinsried/Germany) within the cDNA  
 sequencing consortium of the German Genome Project. This clone  
 (DKFP451J159) is available at the RZPD in Berlin. Please contact  
 the RZPD, Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY. Email: clone@rzpd.de Further  
 information about the clone and the sequencing project is available  
 at <http://mips.gsf.de/proj/cDNA/>.  
 FEATURES  
 source 1..5608 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="RZPD:DKFP451J159"  
 /db\_xref="taxon:9606"  
 /clone="DKFP451J159"  
 /tissue\_type="human skeletal muscle"  
 /clone\_id="451 (synonym: blec1) Vector pSport1; host  
 DH10B; sites NotI + SalI"  
 /dev\_stage="adult"  
 polyA\_signal 5564..5569  
 polyA\_site 5588  
 ORIGIN  
 Alignment Scores: 1.54e-147 Length 5608  
 Pred. No.: 3765.00 Matches 725  
 Score: 3765.00  
 Percent Similarity 99.45% Conservative: 0

Qy 231 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
 Db 1072 AACACGCTACCTCCGTAGGGGCCCCAGGGTCCAGGAGGAGGTGGCTGGCGAGTTTTC 1131  
 Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
 Db 1132 ATCAACAACAGCTTGAACACTCTCAACCGGATGGTCCATTCTCTTCAGGGCGCCATATG 1191  
 Qy 321 ValGluIleSerThrProValLeuIleSerSerSerAsnProSerValIleThrGlnPro 340  
 Db 1192 GTAGAGATCAGCACCCACCTCATCAGCTCCAGCAACCCCTCTGTGATCATCCAGCCCC 1251  
 Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
 Db 1252 ATGGAGAAAGCAGACGTTCTCTTCAGCTGTGGGACAGGTACAGCACTTATCACCCCGCA 1311  
 Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnGlnHisLeu 380  
 Db 1312 CCTGTCTCTCCAGGACATCCACAGCCGTGGTCACTCTCGCTGGCTCCAGCAACACCTC 1371  
 Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
 Db 1372 TCAGCGAACATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCGATGAGCTGGAC 1431  
 Qy 401 LeuGlnLysGlyGluGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
 Db 1432 CTGCAAAAGGAGAGGCGGTACAGGGTCTCTGGGGAAGTGCAGGACGGCTGGCTCAGGGGC 1491  
 Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
 Db 1492 GTCTCTTGTGTCACCGGCGAGTCGGCATCTTCCCAACAAATTACGTCACTCCCATTTTC 1551  
 Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuLysThrThrTrpThrLeu 460  
 Db 1552 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCACATGGACGTTA 1611  
 Qy 461 SerThrSerSerValSerSerGlnGlySerLeuGlyAspProArgGlnSerArg 480  
 Db 1612 TCACCTCTCTGTGTCTCTCCCAAGCAGCATTTTCAGAAAGGTGATCCAGCGCAAGCCGT 1671  
 Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
 Db 1672 CCTTCAAAATCCGTCTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAAAGCAGCCGCGC 1731  
 Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
 Db 1732 CTGGGACTTTAGGACAAAGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGATGGA 1791  
 Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
 Db 1792 TCCCTGCAGAGACCCCTCCAGTCCGGGATCCCACTCTCTGTGGTAGGCTCCCTCAGACGC 1851  
 Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
 Db 1852 AGCCCCACCATGGTCTCTGGCGCTCAGCAGTTCCTCAATCTACAGCCACAGGGGATCCCC 1911  
 Qy 561 SerSerProSerAlaValValGluMetGlySerLysProProAlaLeuThrGlyGluPro 580  
 Db 1912 TCTCTCCCTCAGCCGCTGGTGGAGATGGGGTCCAAAGCCTGCCCTCAGGGGGAGCCCC 1971  
 Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerSerLeu 600  
 Db 1972 GCCCTCAGCTGCATCAGCAGGGGCGAGTGAAGGCCCGGACCCACTCCGGGGCGCAGCTCC 2031  
 Qy 601 IleMetGluAspLysGluIleProlleLysSerGlnProLeuProLysProAlaSer 620  
 Db 2032 ATTAGAAGACAAAGAAATCCCCATCAAGAGTGGAGCTCTGCCAAAAACCCCGCATCT 2091

Best Local Similarity:	99.45%	Mismatches:	4
Query Match:	99.31%	Indels:	0
DB:	9	Gaps:	0
US-1061-201-3 (1-729) x HSM804610 (1-5608)			
Qy	1	MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp	20
Db	53	ATGGATGATTGGACGTTACTTGATCTCTGGAGTGCCCTGTGTGCTTTGAGAAAGCTCGAT	112
Qy	21	ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGluArgVal	40
Db	113	GTACACAGCCAAAGTCTCCCTTGGCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT	172
Qy	41	PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle	60
Db	173	TTCAAAGGCCCAAAAGAGCGTCGGTGGCCCGGAATGCAGGACGGCTGTGTTTTCCAACATT	232
Qy	61	GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln	80
Db	233	GAGGCGCTGCCGGCCAACTGCTGCTCGTGGCCCTCTGGATGGAGTGGCTCAGGGCAG	292
Qy	81	SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg	100
Db	293	AGCTCCGGGAGAGGGGGCTCTTCGCAGGCTGGCACAGATGACCTTGCAGGATGCCAGG	352
Qy	101	LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal	120
Db	353	AAAAGCAGGACCAACCCACAGAGCTTCGACGGCCAGTCTTTCGGCTAGTGCCTAATGTC	412
Qy	121	ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn	140
Db	413	AGAATCCACATGGATGGGGTGCTCGACGCAAAAGGCCTTATGCAACTACAGAGGCAGAAT	472
Qy	141	ProGlyAspLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu	160
Db	473	CCCGGTGACCTAAGGTTTAATAAGGGAGATATCATCTTCTCCGGAGACAGCTGATGAG	532
Qy	161	AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu	180
Db	533	AATTGGTACCAGGGGGAATCAATGGCATCAGCGGGAACCTTCCAGCCAGCTCCGTGGAA	592
Qy	181	ValIleLysGlnLeuProGlnProProLeuProLeuCysArgAlaLeuTyrAsnPheAspLeu	200
Db	593	GTATCATCAAGCAGCTGCCCCAGCGCCCCCGCTCTGCGAGGCCCTCTACAACTTCGACCTA	652
Qy	201	ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle	220
Db	653	CGAGGCCAAGACAAAGATGAGAACCCAGGATTGCTGTGACCTTCCTCAAGGACGATATCATC	712
Qy	221	ThrValLeuSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly	240
Db	713	ACTGTGATACAGCCGAGTGGATGAGAACTGGGCAGAAAGCAAGTTAGGAGATAAGTAGGC	772
Qy	241	IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn	260
Db	773	ATCTTCCCTATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAGAC	832
Qy	261	LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly	280
Db	833	AAAGGTGCGCATCTCTCGCACAAAAAACCTGTCCCTGGTGTCTCTCGTCCAGAGGC	892
Qy	281	AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer	300
Db	893	AACACGCTTACCTCCCTAGGGGCCAGGGTCCAGGAAAGGTGGCTGGCGAGTTTTTCC	952
Qy	301	IleThrAlaLeuAsnThrLeuLeuAsnArgMetValHisSerProSerGlyArgHisMet	320
Db	953	ATCACAACAGCCTTGACACACTCTCAACCGGATGGTCCATTCTCTTTCAGGGCGGCATATG	1012

661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaSer 680

681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700

683 ACCCCCGGAAAGCCTGAACAGCCAGCCACCTCAAGCGGTCCAGCGTCC 2092

684 TGGGCCCCAGAGTACCGTCTATTGGCCACCGAAGCGGTCCGACTCCGAGACGAG 2152

701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrThrValSerThrAlaSer 720

702 ACAGACCTCCCGGAGAAAGTCAGCTTTGCCAAGGCCACAAACCTGGTGTCTCACTGCCTCA 2212

721 GlyThrGlnThrValPheProSerLys 729

722 GGCACCGCAGACCGTGTTCCTCCAGCAAA 2239

LOCUS AK074234

DEFINITION Homo sapiens cDNA FLJ23654 fis, clone COL10108.

ACCESSION AK074234

VERSION AK074234.1 GI:18676780

KEYWORDS oligo capping, fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1

AUTHORS Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y.,

Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,

Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

TITLE NEDO human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3019)

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,

Shibahara, T., Tanaka, T. and Nakamura, Y.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,

University of Tokyo, Laboratory of Genome Structure, Human Genome

Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan

(E-mail: fldna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,

Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing;

Research Association for Biotechnology; cDNA library construction,

5'- & 3'-end one pass sequencing; Department of Virology and Human

Genome Center, Institute of Medical Science, University of Tokyo

(partly supported by Science and Technology Agency).

FEATURES

source

1..3019

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="COL10108"

/issue\_type="colon"

/clone\_lib="COL"

/note="cloning vector pMEISFL3"

223..1596

/note="unnamed protein product"

/codon\_start=1

/protein\_id="BAB85025.1"

/db\_xref="GI:18676781"

/translation="MDDLTLDDLECPVCFEKLVDYAKVLPCQHTFCPKLQRFKAH

KELRCEKRTPTFSNIEALPANLLVRLLDGVRSGSSGRGSRFRPGMTMLQDGRKS

RTNPRLOASPRVNPVHMDGVPRAKALCNVYQNGPDLRNRGDIILLRRLDE

NWYGEISGNFNPASSVEYIKQLPPPLCLRALYNFDLRGDKSENQDCLTFLKDD

ITVISRVDENVAEGRGDKYGVIFLVPFNLTARHLEKKNRGQSCTKNLSVSS

SSRGNSTLRRGPGSRKYPVGGFSITTAINTLRNMVPSGRHMYEISTPVLISNSP

SVITQPMKEADVPSSCVGVSTYHPAPVSPGHSTA VVSLPQSQQHLSANMFVALHSYS

ORIGIN

Alignment Scores

Pred. No. 101e-147 Length: 3019

Score: 3762.00 Matches: 725

Percent Similarity: 99.73% Conservative: 2

Best Local Similarity: 99.45% Mismatches: 2

Query Match: 99.24% Indels: 0

DB: 9 Gaps: 0

US-10-061-201-3 (1-729) x AK074234 (1-3019)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20

Db 223 ATGGATGATTTTGACGTTACTTTGTGGAGTGGCCCTGTGTGCTTTTGAAGAGCTCGAT 282

Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40

Db 283 GTCAAGCCAAAGATCTCTCTTGGCAGCACACCTTCTGTCAACCATGTCTACAGAGGTT 342

Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60

Db 343 TTCAAGGCCCAACAAAGAGTGGCGGTGCCCGGAATGACGAGCGCTGTGTTTTTCCAACATT 402

Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80

Db 403 GAGCGCTGCGCGGCAACCTGTCTGTCTGCGCCTTCTGGATGGAGTGGCGCTCAGGCAC 462

Qy 81 SerSerGlyArgGlyGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100

Db 463 AGCTCCGGGAGAGGGGGCTCTTCCGCAAGGCCCTGGCAGATGACCTTTCAGGATGGCAGG 522

Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120

Db 523 AAAAGCAGGACCAACCCAGACGTCTCGCGCCAGTCTTTTCGGCTAGTGCCTAATGTC 582

Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140

Db 583 AGAATCCACATGGATGGGTGCTTCGAGCAAGAACGGCCCTTATGCAACTACAGAGGCGCA 642

Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160

Db 643 CCCGGTGACCTAAGGTTTAATAGGGGAGATATCATCTCTTCCGGAGACAGCTTGATGAG 702

Qy 161 AsnTrpTyrGlnGlyGluLeuAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180

Db 703 AATTGGTACCAGGGGGAAATCA GTGGCATCAGCGGGAACCTTCCAGCCAGCTCCCGTGAA 762

Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200

Db 763 GTCAATCAAGCAGCTGCCCGCCCGCCGCTCTCGAGGGCCCTCTACAACCTTCGACCTA 822

Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220

Db 823 CGAGGCAAGGACAAAGAGGTGAGAAACCCAGGATTGGCCGTACCTTCCCTAAGGACGATATCATC 882

Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240

Db 883 ACTGTGATCAGCCCGAGTGGATGAGAACTGGGCAGAGGCAAGTTAGGAGATAAGTAGGC 942

Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260

Db 943 ATCTTCCTCTCTTGTGTAGAGCCAAACCTCACCAGCAACACCTTTTAGAGAGAAAC 1002

Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280

Db 1003 AAAAGGTGCCAGTCATCTCTGCACAAAAAACTGTCCCTGGTGTCTCTCTCCAGAGGC 1062



Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 1063 AACACAGCTACCCCTCCGTAGGGGCCAGGGTCCAGGAGAAAGGTGGCTGGGCAGTTTTC 1122  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 1123 ATCAACAACAGCTTGAACACTCTCAACCGGATGGTCCATTCTCTCTTCAGGGCGCCATATG 1182  
Qy 321 ValGluIleSerThrProValIleIleSerSerAspProSerValIleThrGlnPro 340  
Db 1183 GTAGAGATCAGCACCCCAAGTCTCATCAGCTCCAGCAACCCCTCTGTGATCAACCCAGCCC 1242  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrThrHisProAla 360  
Db 1243 ATGGAGAAAGCAGACGTTCTTCCAGCTGTGTGGACAGGTCAAGCACTTATCAACCCCGCA 1302  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnHisLeu 380  
Db 1303 CCTGTCTCTCAGGACATTCACAGCCGTGGTCAAGTCTGCCTGGCTCCAGCAACACCTC 1362  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerThrAlaHisGlyProAspGluLeuAsp 400  
Db 1363 TCAGCGAACAATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1422  
Qy 401 LeuGlnLysGlyGluGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1423 CTGCAAAAGGGAGAGCGGCTCAGGGTCTCTGGGGAAGTGCCAGGACGGCTGGCTCAGGGGC 1482  
Qy 421 ValSerLeuValThrArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1483 GTCTCTTGGTCAACCGCGGAGTCGGCATCTTCCCAACAATACGTTCATCCCATTTTC 1542  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrTrpThrLeu 460  
Db 1543 AGAAAGACCTCTAGTTTTCAGACTCCCGAGAGCCCTGGTCTCTACACCACATGAACGTTA 1602  
Qy 461 SerThrSerValSerGlnGlySerIleSerGluGlyAspProArgGlnSerArg 480  
Db 1603 TCCACCTCTCTGTCTCTCCCAAGCAGCATTTTCAGAAAGTGATCCACGGCAAGCGGT 1662  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1663 CCTTCAAATCCGCTTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAACAGCCCGGC 1722  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1723 CCTGGGACTTTAGGACAAGGTCTCTTGGGAAAGGCGGAGCAGCATGAGAAAGAAATGGA 1782  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1783 TCCCTGCAAGACCCCTCCAGTCCGGGATCCCACTCTCTGTGGTAGGCTCCCTCAGAGCG 1842  
Qy 541 SerProThrMetValLeuArgProGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1843 AGCCCAACCATGGTCTTGGGCTCAGCATGTTCCAAATCTACAGCCAGCGGGATCCCC 1902  
Qy 561 SerSerProSerAlaValValGluMetGlySerLysProAlaLeuThrGlyGlnPro 580  
Db 1903 TCCCTCCCTCAGCCCGTGGTGGAGATGGGGTCCAAAGCTCGCCCTCAGGGGGAGGCC 1962  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1963 GCCCTCAGTCATCAGCAGGGGAGTGAGGCCCGGATCCCACTCCCGGGCCAGCTCCCTC 2022  
Qy 601 IleMetGluAspLysGluIleProIleLysSerGluProLysProProAlaSer 620  
Db 2023 ATTATGGAGACAAAGAAATCCCCATCAAGAGTGAAGCTCTGTGCCAAAAACCGCCGCACT 2082

Qy 621 AlaProProSerLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
Db 2083 GCCCCACCATCTCTGGTGAACCCAGAAAACTCAAGAAATGGCATCGAAAAAGCAAGTC 2142  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro 660  
Db 2143 AAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACCAAAACATTACACCTCCCATCC 2202  
Qy 661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
Db 2203 ACCTCCGGAAGAGCTGAACAGCCAGCCACCCCTCAAGGCGTCCAGGCTGAAAGCAGGCTCC 2262  
Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2263 TTGGGCCAGAGATGACCGTCTATTGTGCCACCGAAGTGGCTGCCACTCCGGACAGCAG 2322  
Qy 701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrLeuValSerThrAlaSer 720  
Db 2323 ACAGACCTCCGGAGAAAGTCAGCTCTTGCCAGGCCACACCCCTGGTGTCCACTGCCCTCA 2382  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
Db 2383 GGCACGCAGACCGTGTTCACGCAAA 2409

RESULT 5  
AX530497  
LOCUS AX530497 1855 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 5 from Patent EPI239051.  
ACCESSION AX530497  
VERSION AX530497.1 GI:25252377  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Shannon.M.  
TITLE Human post-like protein 1  
JOURNAL Patent: EP 1239051-A 6 11-SEP-2002;  
Aeomica, Inc (US)  
FEATURES  
Location/Qualifiers  
source 1. 1855  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,08e-125 Length: 1855  
Score: 3219.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.91% Indels: 0  
DB: 6 Gaps: 0  
US-10-061-201-3 (1-729) x AX530497 (1-1855)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTGACCGTTACTTGATCTCTCTGGAGTGGCCCTGTGCTTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTCAAGCCAAAGTCTCTCTGGCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAGGCCCCAACAAAGAGTGGCGTGGCCCGAATGCGAGGAGCGCTGTGTTTCCACATT 180

Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuValArgLeuLeuValArgSerGlyGln 80  
Db 181 GAGGGCGCTGCGCGGCAACCTGCTGCTGCGCGCTTCTGGATGGAGTGGCTCAGGGCGAG 240  
Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGAGAGGGGGCTCTCCGAGCGCTGGCAGATGACCTTGCAGGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGACCAACCCAGAGCTGCGAGGCCAGTCTTCCGGCTAGTGCCTAATGTC 360  
Qy 121 ArgHisMetAspGlyValProArgAlaLysAlaLysCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGTGCTCGAGCAAGAGGCCTTATGCAACTACAGAGGCGAGAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCCGGTGACCTAAGGTTTAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480  
Qy 161 AsnTyrThrGlnGlyGlnLeuAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTTGGTACCAAGGGGAAATCAATGGCATCAGCGGAAATTTCCAGCCAGCTCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTATCAAGCAGCTGCCCCAGCCCGCCCCCTCTGAGGGCCCTCTACAACTTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGlnAsnGlnAspCysLysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGGCAAGGCAAGAGTGAACAAGGATTCCTGACCTTCCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCGAAGCAAGTTAGGAGATAAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 721 ATCTTCCTCATCTGTTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAC 780  
Qy 261 LysGlyArgGlnSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAGGTCCGCAAGTCACTCTCGCAAAACCTGTCCCTGGTGTCTCTCTCCAGGCG 840  
Qy 281 AsnThrThrLeuArgArgGlyProGlySerArgLysValProGlyGlnPheSer 300  
Db 841 AACACGTCTACCTCCGTAGGGGCCAGGGTCCAGGAGGAGGTGCTGGCGAGTTTCC 960  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAAGCCCTTGAACACTCTCAACCGGATGGTCCATCTCTCTCAGGGCGGCATATG 960  
Qy 321 ValGluLeuSerThrProValLeuLeuSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGCACCCCAAGTCTCATCAGCTCCAGCAACCCCTCTGTGTATCATCCAGCCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGACGCTCTCTCCAGCTGTGTGGGACAGGTCAGCACTTATCACCCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCACAGCCGTGGTGTGCTGCTGCCGTCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAACAATGTTTGTAGCCCTGCACCTCTACTACGCCCATGGACCGCATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlnGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420

Db 1201 CTGCAAAAGGGAGAAAGGCGTCAGGCTCTGGGAAAGTGCCAGGCGCTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTCTTGGTCAACCGGGGAGTGGGCATCTTCCCAAAACAATTACGTATCCCCATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu 460  
Db 1321 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerSerGlnGlySerIleSerGlyAspProArgGlnSerArg 480  
Db 1381 TCCACCTCTCTGTCTCTCCCAAGGCAGCATTTTCAGAAAGGTGATCCACGGCAAGCCGT 1440  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1441 CCCTTCAAAATCCGCTCTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAGCACAGCCGCGC 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1501 CCTGGGACTTTAGGACAAAGGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGAAATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGCAGACACCCCTCCAGTCCGGGATCCCCACTCTCTGGTAGGCTCCCTCAGACGC 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1621 AGCCCCACCATGGTCTCTCCGCTCAGCAGTTCACAAATTTCTACCAAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValGlnMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1681 TCCTCCCTCAGCCGCTGGTGGAGATGGGGTCCAAAGCTGCCCTCAAGGGGAGGCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1741 GCCCTCAGCTGCATCAGCAGGGGCGAGTGAGGCCCGGACCCCACTCCCGCGGCCACGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProlLeuLysSerGluProLeuProLysProPro 618  
Db 1801 ATTATGGAAGACAAAGAAATCCCCATCAGAGTGAGCTCTGCCAAACCGCCC 1854

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 7, 2004 08:02:08 Search time 641 Seconds  
(without alignments)  
4831.414 Million cell updates/sec

Title: US-10-061-201-3

Perfect score: 3791

Sequence: 1 MDDILLLDLLECPVCFEKL.....AKAATLVSTASGTQTVPPSK 729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 3373863 seqs. 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match: 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xip  
-Q=/cgn2\_1/USPTO\_spool\_p/US10061201/runat\_05102004\_174113\_9264/app\_query.fasta\_1.903  
-DB=N\_Geneseq\_29Jan04\_QFMT=fastap-SUFFIX=mg-MINMATCH=0-1-LOOPCL=0  
-LOOPEXT=0-UNITS=bis-START=1-END=1-MATRIX=blosum62-TRANS=human40-cdi  
-LIST=45-DOCALLIGN=200-THR\_SCORE=opt-THR\_MAX=100-THR\_MIN=0-ALIGN=15  
-MODE=LOCAL-OUTFMT=opt-NORM=ext-HEADSIZE=500-MINLEN=0-MAXLEN=2000000000  
-USER=US10061201\_@CGN1\_1\_208\_@runat\_05102004\_174113\_9264-NCPU=6-ICPU=3  
-NO\_MMAPP-LARGEQUERY-NEO\_SCORES=0-WAIT\_DSPBLOCK=100-LONGLOG  
-DEV\_TIMEOUT=120-WARN\_TIMEOUT=30-THREADS=10-XGAPOP=10-XGAPEXT=0.5-FGAPOP=6  
-FGAPEXT=7-YGAPOP=10-YGAPEXT=0.5-DELOP=6-DELEXT=7

Database: N\_Geneseq\_29Jan04.\*

- 1: geneseq1980s.\*
- 2: geneseq1990s.\*
- 3: geneseq2000s.\*
- 4: geneseq2001as.\*
- 5: geneseq2001bs.\*
- 6: geneseq2002s.\*
- 7: geneseq2003as.\*
- 8: geneseq2003bs.\*
- 9: geneseq2003cs.\*
- 10: geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description	
1	3791	100.0	2190	6	ABV89291	Abv89291 Human POS	
2	3791	100.0	2497	6	ABV89290	Abv89290 Human POS	
3	3219	84.9	1855	6	ABV89294	Abv89294 Human POS	
4	3219	84.9	2086	6	ABV89292	Abv89292 Human POS	
5	1409	5	37.2	1008	5	AAH64929	Aah64929 Human sec
6	1098	29.0	766	4	AAL01468	Aal01468 Human rep	
7	1098	29.0	766	4	ABL96913	Abv6913 Human tes	
8	1049	5	27.7	2667	8	ADB80966	Adb80966 RING-SH c
9	1049	5	27.7	5128	8	ADB80967	Adb80967 RING-SH c
10	1049	5	27.7	5205	7	ABX34585	Abx34585 Human mdd
11	1041	5	27.5	3251	6	ABN83326	Abn83326 Human cyt
12	1033	5	27.3	2331	8	ADB80968	Adb80968 RING-SH c
13	1033	5	27.3	2332	4	AAH15208	Aah15208 Human cDN
14	1033	5	27.3	2332	8	ADB80970	Adb80970 RING-SH c
15	1006	26.5	3146	7	ABV72517	Abv72517 Nucleotid	
16	839	22.1	677	2	AAZ17679	Aaz17679 Human gen	
17	816	5	21.5	1369	4	ABK43541	Abk43541 DNA encod
18	721	19.0	1701	6	ABK35062	Abk35062 Human cDN	
19	708	5	18.7	3122	4	ABL08547	AbL08547 Drosophi
20	675	17.8	500	6	ABV89306	Abv89306 Human POS	
21	671	17.7	484	6	ABV89296	Abv89296 Human POS	
22	651	17.2	793	5	AA569854	Aas69854 DNA encod	
c 23	609	5	16.1	5344	4	ABL08546	AbL08546 Drosophi
24	601	15.9	359	6	ABV89303	Abv89303 Human POS	

25	596	15.7	500	6	ABV89313	Abv89313 Human POS
26	595	15.7	359	4	AAI25317	Aai25317 Probe #15
27	595	15.7	359	4	ABA71252	Aba71252 Human fce
28	595	15.7	359	4	AAI51490	Aai51490 Probe #20
29	595	15.7	359	4	ABA37552	Aba37552 Probe #16
30	595	15.7	359	4	AAK45549	Aak45549 Human bon
31	595	15.7	359	4	AAK19553	Aak19553 Human bra
32	595	15.7	359	4	ABS45237	Abs45237 Human liv
33	595	15.7	359	6	ABS19820	Abs19820 Human gen
34	581	15.3	480	3	AAAC08830	Aac08830 Human sec
35	531	14.0	315	6	ABV89299	Abv89299 Human POS
36	528	13.9	500	6	ABV89309	Abv89309 Human POS
37	522	13.8	300	2	AAZ12879	Aaz12879 Human gen
38	509	13.4	480	4	ABS33308	Abs33308 Human liv
39	509	13.4	480	6	ABS08393	Abs08393 Human gen
40	503	13.3	300	2	AAZ12878	Aaz12878 Human gen
41	486	12.8	270	6	ABV89297	Abv89297 Human POS
42	486	12.8	500	6	ABV89307	Abv89307 Human POS
c 43	473	12.5	2241	5	AA572846	Aas72846 DNA encod
44	472	12.5	342	6	ABV89304	Abv89304 Human POS
45	472	12.5	500	6	ABV89314	Abv89314 Human POS

#### ALIGNMENTS

RESULT 1  
ABV89291  
ID ABV89291 standard; cDNA, 2190 BP.  
XX  
AC ABV89291;  
XX  
DT 23-DEC-2002 (first entry)  
XX  
DE Human POSHL1 encoding cDNA ORF SEQ ID NO 2.  
XX  
KW Human; POSHL 1; SH3 domain; POSH-like signalling protein 1; oncogene;  
KW Rho GTPase; signal transduction; gene expression; cancer, vaccine,  
KW gene therapy; transgenic; gene; ss.  
XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT CDS 1..2190  
FT /tag= a  
FT /product= "POSHL1"  
XX  
PN EP1239051-A2.  
XX  
PD 11-SEP-2002.  
XX  
PF 28-JAN-2002, 2002EP-00001165.  
XX  
PR 30-JAN-2001; 2001WO-US000663.  
PR 30-JAN-2001; 2001WO-US000664.  
PR 30-JAN-2001; 2001WO-US000665.  
PR 30-JAN-2001; 2001WO-US000666.  
PR 30-JAN-2001; 2001WO-US000667.  
PR 30-JAN-2001; 2001WO-US000668.  
PR 30-JAN-2001; 2001WO-US000669.  
PR 30-JAN-2001; 2001WO-US000670.  
PR 23-MAY-2001; 2001US-00864761.  
PR 10-OCT-2001; 2001US-0328205P.

XX  
PA (AEOM-) AEOMICA INC.  
XX  
PI Shannon M;  
XX  
DR WPI, 2002-684061/74.  
DR P-PSDB; ABB83999.

XX PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.

XX PS Claim 1, Fig 3, 60pp + Sequence Listing, English.

XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (S1, ABB83959), a sequence having 65% sequence identity to (S1),  
CC (S1) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids.  
CC Human POSHL1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway. (I) is useful  
CC for identifying a specific binding partner. (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1, including diagnosing and  
CC treating cancer, they useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL1 cDNA sequence of the  
CC invention

XX SQ Sequence 2190 BP, 517 A, 701 C, 553 G, 419 T, 0 U, 0 Other.

Alignment Scores:

Pred No.: 1,7e-206 Length: 2190  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-061-201-3 (1-729) x ABY89291 (1-2190)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTGACGTTACTTCTGGAGTGGCCCTGTGTGCTTTTGGAGAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTACAGCCCAAGTCTCCCTTGGCCAGCACACCTTCTGCCAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAAGGCCCAACAAAGAGCTGCGGTGCCCGAATGCAAGACGCGCTGTGTTTTCACACATT 180  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGGGCGCTGCCGGCCCAACCTGCTGCTGGCCCTTCTGGATGGAGTGGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlyGlySerArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGGAGAGGGGGCTCTCCGACGGCTGGCACATGACCTTGGCAGGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAAGCAGACCAACCCCAAGCTGCTGCAAGCCAGCTGCTCCGGCAGTGCCTAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGTGCTCGACGCAAAAGGCCTTATGCAACTACAGAGGGCAGAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCCGGTGACCTAAGGTTTAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480

Qy 161 AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTGGTACCAGCGGGAAATCAATGGCATCAGCGGAACTTCCAGCAGCTCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCAATCAAGCAGCTGCCCCAGCGCGCCCGCTCTGCAGGGCCCTCTACAATCTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGCACAAGGACAAGAGTGGAGAACAGGATTGCGTACCTTCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAGAAAGGCAAGTTAGGAGATAAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGlnLysAsn 260  
Db 721 ATCTTCTCTATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAGAAC 780  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAGGTCCCGCAGTCACTCTGCACAAAAACCTGTCTCCCTGGTGTCTCTCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgLysValProGlyGlnPheSer 300  
Db 841 AACACGCTACTACCTCTCCGTAGGGGCCAGGGTCCAGAGGAAGGTGCTGGGCAGTTTTC 900  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAACAGCTTGAACACTCTCAACCGGATGGTCCATTCTCTTCAGGGCGGCATATG 960  
Qy 321 ValGluIleThrProValLeuIleSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGACCCCACTGCTCATCAGCTCCAGCAACCCCTCTGTGTATCATCCACGCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGAGCGTTCTCTCCAGCTGTGTGGGACAGGTACAGCATTACCCCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCCACAGCCGTGGTCACTGTGCTGCCGTCCCAAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAAACATGTTTGTAGCCCTGCACTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGAGAGAGGGCTCAGGGTCTGGGGAAGTGCCAGGACGGCTGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTTGTGTCACCGGGGCGAGTGGGATCTTCCCAAAACAATTACGTATCCCCATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrTrpLeu 460  
Db 1321 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCACATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerSerGlnHisSerGlnHisGlyAspProArgGlnSerArg 480  
Db 1381 TCCAGCTCTCTGTGCTCTCCCAAGCAGCATTTTCAGAAAGGTGATCCACGGCAAGCCGT 1440  
Qy 481 ProPheLysSerValPheValProThrThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1441 CCCTTCAAAATCCGCTCTTTGTGCCACTGCCATATGCAACCCCGTGAGAAAGCACAGCCGCG 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520

Db 1501 CCTGGGACTTTAGGCAAGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGAAATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGACAGAGACCCCTCCAGTCGGGGATCCCACTCTCGTGGTATAGGCTCCTCAGAGCG 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1621 AGCCCCACCATGGTCTCTCGGCTCAGCAGTTCACATCTACCAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValValGluMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1681 TCTCTCCCTCAGCCCTGGTGGTGGAGATGGGGTCCAGCCCTGCCCCCTCACGGGGGAGGCC 1740  
Qy 581 AlaLeuThrCysLeSerArgGlySerGluAlaArgThrHisSerAlaAlaSerLeu 600  
Db 1741 GCCCTCAGTCATCAGCAGGGGAGTGAGGCCCGGACCCACTCCGCGGCCAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProLysSerGluProLeuProLysProAlaSer 620  
Db 1801 ATTATGGAGAGACAAAGAAATCCCCATCAAGAGTGAAGCTCTGCCCCAAAACCCCGGCATCT 1860  
Qy 621 AlaProProSerLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
Db 1861 GCCCCACCATCCATCTCTGGTGAAACACAGAAATCTCAAGAAATGGCATCGAAAGCAAGTC 1920  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrThrLysHisTyrThrSerHisPro 660  
Db 1921 AAAACCGTGAGATTTCAGAAATTACAGCCCTCTCCCAACCAACATTACACCTCCCATCCC 1980  
Qy 661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
Db 1981 ACCTCCGGAAGCCCTGAACAGCCAGCCACCCCTCAAGGGCGTCCAGCGCTGAAGCAGCGTCC 2040  
Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2041 TTGGGGCCAGAGATGACCGCTCTATTTGGCCCAACCGAAGTGCGCTGCCACTCCGGACAGCAG 2100  
Qy 701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrThrLeuValSerThrAlaSer 720  
Db 2101 ACAGACCTCCGGAGAAAGTCAGCTCTTGCCAAAGGCCACAAACCCCTGGTGCTCCACTGCCTCA 2160  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
Db 2161 GGCACGCGAGACCGTGTTCGCCAGCAAA 2187  
RESULT 2  
ABV89290  
ID ABV89290 standard: cDNA, 2497 BP.  
XX AC ABV89290,  
XX DT 23-DEC-2002 (first entry)  
XX DE Human POSHL1 full length encoding cDNA SEQ ID NO 1.  
KW Human; POSHL1; SH3 domain; POSH-like signalling protein 1; oncogene;  
KW Rho GTPase; signal transduction, gene expression, cancer, vasculature;  
KW gene therapy; transgenic; gene; ss  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX FT CDS 232..2421  
XX FT /tag= a "POSHL1"  
XX FT /product= "POSHL1"  
XX PN EP1239031-A2

XX PD 11-SEP-2002.  
XX PF 28-JAN-2002; 2002EP-00001165.  
XX PR 30-JAN-2001; 2001WO-US0000663.  
XX PR 30-JAN-2001; 2001WO-US0000664.  
XX PR 30-JAN-2001; 2001WO-US0000665.  
XX PR 30-JAN-2001; 2001WO-US0000666.  
XX PR 30-JAN-2001; 2001WO-US0000667.  
XX PR 30-JAN-2001; 2001WO-US0000668.  
XX PR 30-JAN-2001; 2001WO-US0000669.  
XX PR 30-JAN-2001; 2001WO-US0000670.  
XX PR 23-MAY-2001; 2001US-00864761.  
XX PR 10-OCT-2001; 2001US-0328205P.  
XX PA (AEOM-) AEOMICA INC.  
XX PI Shannon M.  
XX DR WPI, 2002-684061/74.  
XX DR P-PSDB; ABB83999.  
XX PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL.  
XX PT -1, useful for treating disorders associated with decreased expression or  
XX PT activity of human POSHL1.  
XX PS Claim 1; Fig 3; 50pp + Sequence Listing; English.  
XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
XX CC protein 1 (POSHL 1) polypeptide (I), comprising a sequence of 730 amino  
XX CC acids (SI, ABB83999), a sequence having 65% sequence identity to (SI),  
XX CC (SI) having 95% deviations, especially conservative substitutions or a  
XX CC fragment of the sequences comprising at least 8 contiguous amino acids.  
XX CC Human POSHL 1 is a proto-oncogene/oncogene product that functions as an  
XX CC adaptor protein that interacts with Rho family small GTPases as well as  
XX CC downstream components of the signal transduction pathway. (I) is useful  
XX CC for identifying a specific binding partner (I) and nucleic acids (II)  
XX CC encoding (I) are useful for diagnosing, monitoring disease and treating  
XX CC caused by altered expression of human POSHL1 including diagnosing and  
XX CC treating cancer, they are useful in the development of vaccines and (II) is  
XX CC useful in gene therapy. (II) is useful for constructing microarrays which  
XX CC are useful for measuring and for surveying gene expression and creating  
XX CC transgenic non-human animals capable of producing the proteins. The  
XX CC present sequence is that of a human POSHL 1 cDNA sequence of the  
XX CC invention  
SQ Sequence 2497 BP; 603 A; 777 C; 638 G; 479 T; 0 U; 0 Other;  
Alignment Scores  
Prod No: 1,99e-206 Length: 2497  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps 0  
US-10-061-201-3 (1-729) x ABV89290 (1-2497)  
Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 232 ATGGATGATTTTGACGTTACTTCTCTGGAGTGGCCCTGTGTGCTTTGAGAGCTCGAT 291  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 292 GTCACAGCCAAAGTCTCTCCCTGCCAGCAGCACCTTCTGCAAAACCATGTCTACAGAGGTT 351  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60



RESULT 3  
 ABV89294  
 ID ABV89294 standard; cDNA, 1855 BP.  
 XX AC ABV89294.  
 XX DT 23-DEC-2002 (first entry)  
 XX DE Human POSHL1 partial encoding polynucleotide SEQ ID NO 6.  
 XX KW Human; POSHL1; SH3 domain; POSH-like signalling protein 1; oncogene;  
 XX KW Rho GTPase; signal transduction, gene expression, cancer, vaccine,  
 XX KW gene therapy, transgenic, ss.  
 XX OS Homo sapiens.  
 XX PN EP1239051-A2.  
 XX PD 11-SEP-2002.  
 XX PF 28-JAN-2002, 2002EP-00001165.  
 XX PR 30-JAN-2001; 2001WO-US000663.  
 XX PR 30-JAN-2001; 2001WO-US000664.  
 XX PR 30-JAN-2001; 2001WO-US000665.  
 XX PR 30-JAN-2001; 2001WO-US000666.  
 XX PR 30-JAN-2001; 2001WO-US000667.  
 XX PR 30-JAN-2001; 2001WO-US000668.  
 XX PR 30-JAN-2001; 2001WO-US000669.  
 XX PR 30-JAN-2001; 2001WO-US000670.  
 XX PR 23-MAY-2001; 2001US-00864761.  
 XX PR 10-OCT-2001; 2001US-0328205P.  
 XX PA (AEOMC-) AEOMICA INC.  
 XX PI Shannon M.  
 XX DR WPI, 2002-684061/74.  
 XX PT Novel human SH3 domain (POSH)-like signalling protein 1 polypeptide, POSHL  
 PT -1, useful for treating disorders associated with decreased expression or  
 PT activity of human POSHL1.  
 XX PS Example 1; SEQ ID NO 6, 60pp + Sequence Listing, English.  
 CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
 CC protein 1 (POSHL1) polypeptide (I), comprising a sequence of 730 amino  
 CC acids (S1, ABB83599), a sequence having 65% sequence identity to (S1),  
 CC (S1) having 95% deviations, especially conservative substitutions or a  
 CC fragment of the sequences comprising at least 8 contiguous amino acids  
 CC Human POSHL1 is a proto-oncogene/oncogene product that functions as an  
 CC adaptor protein that interacts with Rho family small GTPases as well as  
 CC downstream components of the signal transduction pathway (II) is useful  
 CC for identifying a specific binding partner. (I) and nucleic acids (II)  
 CC encoding (I) are useful for diagnosing, monitoring disease and treating  
 CC caused by altered expression of human POSHL1 including diagnosing and  
 CC treating cancer; they are useful in the development of vaccines and (II) is  
 CC useful in gene therapy. (II) is useful for constructing microarrays which  
 CC are useful for measuring and for surveying gene expression and creating  
 CC transgenic non-human animals capable of producing the proteins. The  
 CC present sequence is that of a human POSHL1 cDNA sequence of the  
 CC invention. Note: The present sequence did not form part of the printed  
 CC specification, but is based on sequence information supplied to Derwent  
 CC by the European Patent Office  
 XX SQ Sequence 1855 BP, 424 A; 581 C; 485 G; 365 T; 0 U; 0 Other.

Alignment Scores:

Pred No.: 4.7e-174 Length: 1855

Score: 3219.00 Matches: 618  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 84.91% Indels: 0  
 DB: 6 Gaps: 0  
 US-10-061-201-3 (1-729) x ABV89294 (1-1855)  
 Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
 Db ! ATGGATGATTTTGACGTTACTTGATCTCTGGAGTGCCTCTGTGCTTTGAGAAAGCTCGAT 60  
 Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
 Db 61 GTCACAGCCAAAGTCTCTCCCTTGGCAGCAGCACACCTTCTGCAAAACCATGCTACAGAGGGTT 120  
 Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
 Db 121 TTCAAGGCCCCACAAAGAGCTGCGGTGCCCGGAATGCAGGACGCTGTGTTTCCAACATT 180  
 Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
 Db 181 GAGGCGCTGCCGGCCAACTGTGCTGTGCGCCCTCTGGATGGAGTGGCTCAGGGCAG 240  
 Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
 Db 241 AGCTCCGGGAGAGGGGCTCTCTCCCGCAGGCCCTGGCAGCATGACCTTCAGGATGGCAGG 300  
 Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
 Db 301 AAAAGCAGGACCAACCCAGAGCTCTGCAGGCCAGTCTTCCGGCTAGTGCCTAAATGTC 360  
 Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
 Db 361 AGAATCCACATGGATGGGTGCCCTCGAGCAAGGCCCTTATGCAACTACAGAGGCGAGAAT 420  
 Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu 160  
 Db 421 CCCCGTGACCTTAAGGTTTAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480  
 Qy 161 AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
 Db 481 AATTGGTACCAGGGGAAATCAATGGCATCAGCGGAACTTCCAGCCAGCTCCCGTGAA 540  
 Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
 Db 541 GTCATCAAGCAGCTGCCCGCCCGCCGCTCTGCAGGGCCCTCTACAACTTCGACCTA 600  
 Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220  
 Db 601 CGAGGCAAGGACAAGAGTGAAGAACCAAGGATTGCGCTGACCTTCCTCAAGGACGATATCATC 660  
 Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
 Db 661 ACTGTGTATCAGCCGCGGTGGATGAGAACTGGGCGCAAGAGCAAGTTAGAGATAAAGTAGGC 720  
 Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
 Db 721 ATCTTCCCTATCTGTTTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAGAAC 780  
 Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerSerArgGly 280  
 Db 781 AAAAGTCCGACAGTCATCTCTGCACAAAAAACCTGTCTCCCTGGTGTCTCTCTCCAGAGGC 840  
 Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
 Db 841 AACACGCTACCTCCCTAGGGGCCCGAGGGTCCAGGAGGAAGGTGCTCGGCGCATTTTCC 900  
 Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320

DT 23-DEC-2002 (first entry)  
XX DE Human POSHL1 encoding cDNA nucleotides 1-2086 SEQ ID NO 4.  
XX KW Human; POSHL1, SH3 domain; POSH-like signalling protein 1; oncogene;  
KW Rho GTPase; signal transduction; gene expression; cancer; vaccine;  
XX KW gene therapy; transgenic; ss  
XX OS Homo sapiens.  
XX PN EPI239051-A2.  
XX PD 11-SEP-2002.  
XX PF 28-JAN-2002; 2002EP-00001165.  
XX PR 30-JAN-2001; 2001WO-US0000663.  
XX PR 30-JAN-2001; 2001WO-US0000664.  
XX PR 30-JAN-2001; 2001WO-US0000665.  
XX PR 30-JAN-2001; 2001WO-US0000666.  
XX PR 30-JAN-2001; 2001WO-US0000667.  
XX PR 30-JAN-2001; 2001WO-US0000668.  
XX PR 30-JAN-2001; 2001WO-US0000669.  
XX PR 30-JAN-2001; 2001WO-US0000670.  
XX PR 23-MAY-2001; 2001US-00864761.  
XX PR 10-OCT-2001; 2001US-0328205P.  
XX PA (AEOM-) AEOMICA INC.  
XX PI Shannon M;  
XX DR WPI, 2002-684061/74.  
XX PT Novel human SH3 domain (POSH)-like signalling protein 1 polypeptide, POSHL  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.  
XX PS Claim 4; SEQ ID NO 4; 60pp + Sequence Listing; English.  
XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (S1, ABB83999), a sequence having 65% sequence identity to (S1),  
CC (S1) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids  
CC Human POSHL1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway. (I) is useful  
CC for identifying a specific binding partner (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1 including diagnosing and  
CC treating cancer, they are useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL1 cDNA sequence of the printed  
CC invention. Note: The present sequence did not form part of the printed  
CC specification, but is based on sequence information supplied to Derwent  
XX CC by the European Patent Office  
SQ Sequence 2086 BP; 482 A; 639 C; 558 G; 407 T; 0 U; 0 Other;

Alignment Scores:  
Pred No.: 539e-174 Length: 2086  
Score: 3219.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.91% Indels: 0  
DB: 6 Gaps: 0

Db 901 ATCAACACGCGCTTGAACACTCTCAACCGGATGGTCCATTCTCTTCAGGCGGCCATATG 960  
Qy 321 ValGluLeuSerThrProValLeuLeuSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGACACCCAGTGCTCATAGCTCCAGCAACCCCTCTGTGATCACCACGCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGACAGCTTCTTCCAGCTGTGTGGACAGGTACAGCACTTATCACCCGCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCACACGCGGTGGTCAGTCTGCCTGGCTCCACCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAACATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlnGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGGAGAAAGCGGTCTCAGGCTCTGGGGAAAGTGCACGAGCGGTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTTGGTCACCGGGCGAGTCGGCATCTTCCCAACAATTACGTCATCCCAATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrTrpThrLeu 460  
Db 1321 AGAAAGACCTCTAGTTTTCCAGACTCCCGGAGCCCTGGTCTCTACACCATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerGlnGlySerIleSerGluGlyAspProArgGlnSerArg 480  
Db 1381 TCACCTCTCTGTCTCTCCCAAGGCAGCATTTTCAGAGGTGATCCACGGCAAGCCGT 1440  
Qy 481 ProPheLysSerValPheValProThrAlaLeuValAsnProValArgSerThrAlaGly 500  
Db 1441 CCTTCAAAATCCGCTTTGTGCCCACTGCCCATAGTCAACCCCGTGAGAGACACGCCCGC 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLysArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1501 CTGGGCACTTTAGGACAAGGGTCTCTTCGAAAGGGCGGAGCAGCATGAGAAAGAAATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGCAGAGACCCCTCCAGTCCGGGATCCCCACTCTCTGTGGTAGGCTCCCTCAGACGC 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1621 AGCCCCACCATGGTCTCTCGGCCTCAGCAGTTTCCAATTCTACAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValGluMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1681 TCTCTCCCCCTCAGCCGTGGTGGTGGAGATGGGGTCCAAAGCCTGCCCTCAGGGGGAGCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1741 GCCCTCACGTGCATCAGCAGGGGCGAGTGAGGCGCCGACCCACTCCGGCGCCAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProLeuLysSerGluProLeuProLysProPro 618  
Db 1801 ATTATGGAAGACAAAGAAATCCCCATCAAGAGTGAGGCTCTGCCAAACCCGCC 1854

RESULT 4  
ID ABV89292  
AC ABV89292 standard; cDNA; 2086 BP.  
XX AC ABV89292,  
XX



Qy	1 MetAspAspLeuThrLeuLeuLeuLeuLeuCysProValCysPheGluLysLeuAsp 20 232 ATGGAGAAAGACAGCGTTCTCCAGCTGTGTGGGACAGGTTCAGCACTTATCACCCCGCA 131
Db	361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnHisLeu 380 1312 CCTGTCTCTCCAGGCATCTCCACAGCCGTGGTGCACTGTCTGGCTGCCAGCAACACCTC 1371
Db	381 SerAlaAsnMetPheValAlaLeuHisHisSerTySerAlaHisGlyProAspGluLeuAsp 400 1372 TCAGCGAACAATGTTGTAGCCCTCGCATCTACTACGCCCATGGACCCGATGAGCTGGAC 1431
Db	401 LeuGlnLysGlyGluGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420 1432 CTGC AAAAGGGAGAGCGGTCAAGGGTCTCTGGGGAAGTGCACGAGCGGCTGGCTCAGGGGC 1491
Qy	421 ValSerLeuValThrGlyArgValGlyLePheProAsnAsnTyrValleProllePhe 440 1492 GTCTCTCTGTGTACC GGCGGAGTGGGCATCTTCCCAAACAATTACGTTCATCCCCATTTTC 1551
Db	441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrThrLeu 460 1552 AGAAAAGACCTCTAGTTTTTCCAGACTCCCGGAGCCCTGGTCTCTACACCATGACGTTA 1611
Db	461 SerThrSerSerValSerSerGlnGlySerLeuGlyAspProArgGlnSerArg 480 1612 TCCACTCTCTGTGTCTCCCAAGGCAGCATTTTCAGAAAGTGTATCCACGGCAAAAGCCGT 1671
Qy	481 ProPheLysSerValPheValProThrAlaLeuValAsnProValArgSerThrAlaGly 500 1672 CCCCTTCAAACTCCGCTTTTGTGGCCACTGCCATAGTCAACCCCCTGGAGAAGCACAGCCGCG 1731
Db	501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520 1732 CTTGGGACCTTAGGACAAGGGTCTCTTCGGAAAGGGGGAGCAGCATGAAAAAGAAATGGA 1791
Qy	521 SerLeuGlnArgProLeuGlnSerGlyLeProThrLeuValValGlySerLeuArgArg 540 1792 TCCCTGCAGAGACCCCTCCAGTCCGGGATCCCCACTCTCTGGTGTAGGCTCCCTCAGACGC 1851
Db	541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyLePro 560 1852 AGCCCAACCATGGTCTCTTCGGCTCAGCAGCATTTCCAATTTCTACACGCCACAGGGGATCCCC 1911
Qy	561 SerSerProSerAlaValValAlaGlnMetGlySerLysProAlaLeuThrGlyGluPro 580 1912 TCCCTCCCTCAGCCGTGGTGGAGTGGGTGCCAAGCCCTGCCCTCAGCGGGAGGCC 1971
Db	581 AlaLeuThrCysIleSerArgGlySerGlyAlaArgThrHisSerAlaAlaSerSerLeu 600 1972 GCCTTCACGTGCATCAGCAGGGGCAAGTGAGGCCCGGACCCCACTCCGCGGCCAGCTCCCTC 2031
Qy	601 IleMetGluAspLysGluIleProlleLysSerGluProLeuProLysProPro 618 2032 ATTATGGAAAGACAAAGAAATCCCCATCAAGAGTGAGGCTCTGCCAAAACCGGCC 2085
Db	total number of his satisfying chosen parameters 55026578
Qy	Minimum DB seq length: 0 Maximum Match 0% Maximum Match 100% Listing first 45 summaries
Db	Post-processing
Qy	Command line parameters: -MODEL=framer- p2n.model -DEV=xlp -q=regnt_1/U/SPTO_spool_pUS1006120/runat_05102004_174114_9285/app_query.fasta_1_903 -DB=EEST-QMFT-fastap-SUFFIXX=rsr-MINMATCH=0.1-LOOPCL=0-LOOPEXT=0

```

MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10061201/runat_05102004_174114_9285/app_query.fasta_1.903
-DB=EST -OEMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

```

-UNITS=bits -START=1 -END=1 -MATRIX=blastm62 -TRANS=human40.cdi -LIST=45  
 -DOALIGN=200 -THR\_SCORE=pet -THR\_MIN=100 -THR\_MAX=100 -ALIGN=15 -MODE=LOCAL  
 -OUTFMT=ps -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US10061201 @CGN\_1\_1\_5180 @runat\_05102004\_174114\_9285 -NCPU=6 -ICPU=3  
 -NO\_MMALP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=5 -RGAPOP=6  
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=5 -DELOP=6 -DELEXT=7

Database: EST:

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estm.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_esti.\*
- 9: gb\_est2.\*
- 10: gb\_est3.\*
- 11: gb\_est4.\*
- 12: gb\_est5.\*
- 13: gb\_est6.\*
- 14: gb\_est7.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vit.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vit.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Match	Length	DB	ID	Description
1	3615.5	95.4	2199	29	AY417552	Homo sapi
2	3538.5	93.3	2199	29	AY417553	Pan trogl
3	3088	81.5	3989	11	AK033601	Mus muscu
4	2946.5	77.7	2205	29	AY417554	Mus muscu
5	2891	76.3	4447	11	AK029168	Mus muscu
6	2294	60.5	3182	11	AK028153	Mus muscu
7	1454	38.4	865	13	BQ931816	BO931816 AGENCOURT
8	1348	35.6	1114	12	BM4553267	BM4553267 AGENCOURT
9	1263	33.3	824	12	B1827373	B1827373 603077734
10	1240.5	32.7	931	12	B1519965	B1519965 603077153
11	1236	32.6	802	12	B1827187	B1827187 603077395
12	1236	32.6	972	12	B1520050	B1520050 603071361
13	1202	31.7	1158	10	B615992	B615992 601280026
14	1175	31.0	1159	12	B1825238	B1825238 603071926
15	1169.5	30.8	776	12	B1830329	B1830329 603073063
16	1167	30.8	819	10	BE736034	BE736034 601305637
17	1132.5	29.9	807	12	B1561960	B1561960 603254888
18	1124	29.6	772	12	BG218826	BG218826 RST38570
19	1056	27.9	721	12	B1461060	B1461060 603206926
c 20	1048	27.6	730	12	B1520952	B1520952 603071361
21	1046	27.6	660	10	BE616761	BE616761 601279705

22 1001.5 26.4 3521 11 AK029202  
 23 993 26.2 691 10 BE311579  
 24 982.5 25.9 867 13 BQ947482  
 25 934 24.6 763 12 BG190764  
 26 923 24.3 525 13 BX283463  
 27 896.5 23.6 755 14 CA463900  
 28 890 23.5 839 13 BU936713  
 29 884 23.3 556 10 BQ077029  
 30 867 22.9 800 13 BU938619  
 31 858.5 22.6 2036 11 BC033203  
 32 856 22.6 769 14 CA466470  
 33 852 22.5 593 13 BQ930446  
 c 34 832 21.9 948 12 B1520780  
 35 827.5 21.8 815 14 CA463777  
 36 825 21.8 2406 11 AK088003  
 37 816 21.5 683 14 CB854282  
 38 816 21.5 1098 10 BB610623  
 39 815 21.5 648 10 BB613665  
 40 791 20.9 644 10 BB625180  
 41 761 20.1 656 12 BM837901  
 42 727 19.2 555 12 BM030927  
 c 43 725 19.1 840 13 BU745401  
 44 700 18.5 387 12 BM792505  
 45 693 18.3 559 14 CB274053

## ALIGNMENTS

RESULT 1  
 AY417552  
 LOCUS AY417552 2199 bp DNA linear GSS 12-DEC-2003  
 DEFINITION Homo sapiens HCM6262 gene, VIRTUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.  
 ACCESSION AY417552  
 VERSION AY417552.1 GI:39773512  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2199)  
 AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Cvello,D.R., Lu,F., Murphy,B.,  
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
 gene trios  
 JOURNAL Science 302 (5652), 1960-1963 (2003)  
 PUBMED 14671302  
 REFERENCE 2 (bases 1 to 2199)  
 AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
 Todd,M.A., Tanenbaum,D.M., Cvello,D.R., Lu,F., Murphy,B.,  
 Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
 Adams,M.D. and Cargill,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 COMMENT This sequence was made by sequencing genomic exons and ordering  
 them based on alignment.  
 FEATURES Location/Qualifiers  
 source 1..2199  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 <1..>2199  
 /locus\_tag="HCM6262"  
 gene  
 ORIGIN  
 Alignment Scores:

AK029202 Mus muscu  
 BE311579 601142460  
 BQ947482 AGENCOURT  
 BG190764 RST9839 A  
 BX283463 BX283463  
 CA463900 AGENCOURT  
 BU936713 AGENCOURT  
 BQ077029 226821 MA  
 BU938619 AGENCOURT  
 BC033203 Homo sapi  
 CA466470 AGENCOURT  
 BQ930446 AGENCOURT  
 B1520780 603071753  
 CA463777 AGENCOURT  
 AK088003 Mus muscu  
 CB854282 U1-CF-DU1  
 BB610623 BB610623  
 BB613665 BB613665  
 BB625180 BB625180  
 BM837901 K-EST0114  
 BM030927 495479 MA  
 BU745401 CH1#009 E  
 BM792505 K-EST0072  
 CB274053 mai88405.



Db 1909 AATGGCATCGAAAGCAAGTCAAAACCGTGGAGATTTCAGAAATTACAGCCCTCTCTCCACC 1968  
Qy 654 LysHisTyrThrSerHisProThrSerGlyLysProGluGlnProAlaThrLeuLysAla 673  
Db 1969 AAACATTACACTCCCATCCCACTCCGGAAGCCCTGGAACAGCCACCCCTCAAGGGCG 2028  
Qy 674 SerGlnProGluAlaAlaSerLeuGlyProGluMetThrValLeuPheAlaHisArgSer 693  
Db 2029 TCCCAAGCTGAAGCAGCGTCTTGGGCCAGAGATGACCGTCTATTTGGCCACCGAAGT 2088  
Qy 694 GlyCysHisSerGlyGlnGlnThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThr 713  
Db 2089 GGCTGCCACTCCGGACAGCAGACACCTCCGGAGAAAGTCAGCTCTTGGCAAGGCCACA 2148  
Qy 714 ThrLeuValSerThrAlaSerGlyThrGlnThrValPheProSerLys 729  
Db 2149 ACCCTGGTGTCCACTGCTCAGGCACGACCGTGTTCGCCAGCAAA 2196  
RESULT 2  
AY417553  
LOCUS AY417553 2199 bp DNA linear GSS 12-DEC-2003  
DEFINITION Pan troglodytes HCM6262 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY417553  
VERSION AY417553.1 GI:39773513  
KEYWORDS GSS  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
REFERENCE 1 (bases 1 to 2199)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
Todd,M.A., Tenebaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 2199)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
Todd,M.A., Tenebaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES Location/Qualifiers  
source 1..2199  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
gene <1..>2199  
/locus\_tag="HCM6262"  
ORIGIN  
Alignment Scores:  
Pred No.: 1.02e-249 Length: 2199  
Score: 3538.50 Matches: 691  
Percent Similarity: 94.16% Conservative: 2  
Best Local Similarity: 93.89% Mismatches: 12  
Query Match: 93.34% Indels: 11  
DB: 29 Gaps: 2  
US-10-061-201-3 (1-729) x AY417553 (1-2199)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTGACGTTACTTGATCTCTGTGGAGTGCCTGTGTGCTTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTCACAGCCAAAGTAGTCTCCCTTGGCAGCAGACACCTTCTGCAAAACCATGTCTACAGAGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAAGGCCCCACAAAGAGCTGCGGTGCCCGCAATGCAGGACGCTGTGTTTTCACAATT 180  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGCGCTGCGCGCCAAACCTGCTGCTGCTGCGCCCTTCTGGATGGAGTGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGAGAGGGGCTCTTCCGACGGCTGGCACCATGACCTTCGACGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGGACCAACCCCGACAGCTCTCAGGCCAGTCCTTCCGGCTAGTGNNTAAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLysAlaCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGNNNNNNCGAGCAANNNTTATGCAACTACAGAGGGCAGAAT 420  
Qy 141 ProGlyAsnLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCTGGTGCCTAAAGGTTTAATAAGGGAGATATCATCTCTTCCGGAGACACCTTGATGAG 480  
Qy 161 AsnTyrTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTGGTACCAGGGGAAATCAATGGCATCACTGGGAACTTCCAGCCAGCTCCCTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCAATCAAGCAGCTGCCCGAGCGCCCGCTCTCGACGGGCCCTTCAACTTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGCAAGGACAAGAGTGAGAAACAGGATTGGCTGACCTTCCCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCGAAGGCAAGTTAGAGATAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 721 ATCTTCCCTATCTTGTGTAGGCCAAACCTCACCGCAAGACACCTTCTAGAGAAAGAC 780  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAGGTGCGCCAGTCAATCCCGCACAAAACCTGTCCCTGGTGTCTCTCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 841 AACACGTCTACCTCCGTAGGGGCCCGAGGCTCCAGGAGGAAGGTGCTGGGCAATTTTCC 900  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAACAGCCTTGAACAACCTCAACCCGGATGGTCCATTCTCCTTCAGGGCGCCATATG 960  
Qy 321 ValGluIleSerThrProValLeuIleSerSerSerSerSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGCACCCCAAGTGTCTCATCAGTTTCCAGCAACCCCTCTGTGTATCACCAGCCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360

Db 2029 TCCACGGCTGAAGCAGCGTCTTTGGGCCAGAGATGACCGTCTATTTTGGCCACCGAAGC 2088  
Qy 694 GlyCysHisSerGlyGlnThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThr 713  
Db 2089 GGCTGCCACTCCGGACACGACAGACCTCCGGAGAAAAGTCAGCTCTTGGCAAGGCCACA 2148  
Qy 714 ThrLeuValSerThrAlaSerGlyThrGlnThrValPheProSerLys 729  
Db 2149 ACCCTGGTGTCCACTGCCTCAGGCACCGACCGTGTCTTTCCCAAGCAAA 2196

Db 1021 ATGGAAAGCAGACGCTTCTTCCAGCTGTGTGGACAGGTCAAGCACTTATCACCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCACAGCGGTGGTCAGTCTGGCTGCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAACATGTTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGluGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGGAGAGAGCGGTCAAGGTCTCTGGGGAAAGTGCACGAGCGGTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTGGTCACCGGGGAGTCGGCATCTTCCCAACAAATTACGTCAATCCCCATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLysLeuThrTrpThrLeu 460  
Db 1321 AGCACACCTCTACAGTTATCCGGGTGTAGGAGGTGGGGCCA-----ACTTTA 1368  
Qy 461 SerThrSerSerValSerSerGlnGly-----SerIleSerGlu 473  
Db 1369 GTTAAGCCTGTAGNTCACAAAAATGGGTGTGTATGGGTGGGANNNNNGCAGCATTTTCAGAA 1428  
Qy 474 GlyAspProArgGlnSerArgProPheLysSerValPheValProThrAlaIleValAsn 493  
Db 1429 GGTGATCCACAGCAAAAGCGTCCCTTCAATCCGTCTTTGTGCCCACTGCCATAGTCAAC 1488  
Qy 494 ProValArgSerThrAlaGlyProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArg 513  
Db 1489 CCGTGAGAAAGCACAGCGGCCCTGGGACTTTAGGACAAAGGTGNNNTCGGAAAGGCGG 1548  
Qy 514 SerSerMetArgLysAsnGlySerLeuGlnArgProLeuGlnSerGlyIleProThrLeu 533  
Db 1549 AGCAGCATGAGAAAGAAATGGATCCCTGCAGAGACCCCTCCAGTCCGGGATCCCACTCTC 1608  
Qy 534 ValValGlySerLeuArgSerProThrMetValLeuArgProGlnGlnPheGlnPhe 553  
Db 1609 GTGGTAGGCTCCCTCAGACGACGCCCAACCATGGTCTCGGCCCTCAGCAGTTCCAAATTC 1668  
Qy 554 TyrGlnProGlnGlyIleProSerSerProSerAlaValValValGlnMetGlySerLys 573  
Db 1669 TACCAGCCACAGGGGATCCCTCTCCCTCCCTCAGCCANNNNNGTGAGATGGGGTCCAAG 1728  
Qy 574 ProAlaLeuThrGlyGluProAlaLeuThrCysIleSerArgGlySerGluAlaArgThr 593  
Db 1729 CCTGCCCTCAGGGGGAGCCCGCCCTCACGTGCATCAGAGGGGCAGGAGGCCCGGATC 1788  
Qy 594 HisSerAlaAlaSerSerLeuIleMetGluAspLysGluIleProIleLysSerGluPro 613  
Db 1789 CACTCCGGCGCCAGCTCCCTCATTTATGGAAAGACAAAGAAATCCCCATCAAGATGAGCCT 1848  
Qy 614 LeuProLysProProAlaSerAlaProProSerIleLeuValLysProGluAsnSerArg 633  
Db 1849 CTGCCAAAACCCGCCCGCTGCCCAACCATCCATCTCGTGAAACCCAGAAAACTCAAGN 1908  
Qy 634 AsnGlyIleGluLysGlnValLysThrValArgPheGlnAsnTyrSerProProThr 653  
Db 1909 AATGGCATTCGAAAAANCAAGTCAAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACC 1968  
Qy 654 LysHisTyrThrSerHisProThrSerGlyLysProGlnGlnProAlaThrLeuLysAla 673  
Db 1969 AAACATTACACCTCCCATCCCACTCTGGAAAGCCCTGAACAGCCACCCCTCAAGGCG 2028  
Qy 674 SerGlnProGluAlaSerLeuGlyProGluMetThrValLeuPheAlaHisArgSer 693